



OPEN

2D signal estimation for sparse distributed target photon counting data

Matthew Hayman¹✉, Robert A. Stillwell¹, Josh Carnes¹, Grant J. Kirchhoff², Scott M. Spuler¹ & Jeffrey P. Thayer²

In this study, we explore the utilization of penalized likelihood estimation for the analysis of sparse photon counting data obtained from distributed target lidar systems. Specifically, we adapt the Poisson Total Variation processing technique to cater to this application. By assuming a Poisson noise model for the photon count observations, our approach yields denoised estimates of backscatter photon flux and related parameters. This facilitates the processing of raw photon counting signals with exceptionally high temporal and range resolutions (demonstrated here to 50 Hz and 75 cm resolutions), including data acquired through time-correlated single photon counting, without significant sacrifice of resolution. Through examination involving both simulated and real-world 2D atmospheric data, our method consistently demonstrates superior accuracy in signal recovery compared to the conventional histogram-based approach commonly employed in distributed target lidar applications.

Photon counting is employed in a variety of optical detection applications where extremely low light detection is required, or applications leveraging optical quantum information¹. The detection technique has been leveraged in optical communications², medical imaging³, 3D imaging⁴, astronomy⁵ and much more. Photon counting is commonly employed in distributed target (typically atmospheric or bathymetric) lidar where photon fluxes are low. This arises from various factors, including the range of the volume under interrogation, inefficient scattering or collection mechanisms, and limitations imposed by instrument parameters (e.g. low pulse energy, small collection aperture, size, weight, and power restrictions and low efficiency detection processes). In these applications, a crucial step is estimating the amplitude of the backscatter optical signal—a key intermediary of any derived lidar variable where errors in the backscatter estimation problem can only compound and propagate to the derived variables. Traditionally, backscatter estimation is performed by placing observed photon counts into histograms of fixed range and time resolution. The selection of bin widths, and consequently time and range resolutions, are determined based on resolvable signal levels. Larger bin width allows for greater noise suppression but comes at the cost of reduced resolution—a trade-off that must be carefully considered.

The standard approach to processing lidar data treats these histogrammed photon counts as the true backscatter signal. This assumption is typically valid under two conditions: (1) backscatter photon counts in a bin are high, resulting in minimal shot noise, and (2) when the instantaneous photon arrival rate is small throughout the bin interval (negligible nonlinear detector effects are present). In this work, we focus on addressing cases where the first condition (1) is invalid due to significant shot noise. This situation occurs when we aim to observe structure at high resolution, a relative term indicated by the relationship between the capture time of a bin and its associated mean photon flux. Consequently, the absolute photon counts in any given bin becomes relatively small. While the second condition (2) is also important, some issues related to nonlinear effects have been addressed for hard target lidar in previous works^{6,7}. We leave it to future investigations to adapt these approaches for distributed target lidar. However, it should be pointed out that there are ample instances in atmospheric lidar where non-ideal detector behaviors are not significant, particularly for low peak-power lidar systems. For example, the water vapor retrievals obtained from the MicroPulse DIAL (the instrument used for this study) are exclusively limited to linear/ideal detection regimes with significant success^{8,9}.

In placing photon counts in a histogram, the expected number of observed photons α_i in a bin i is given by

¹National Science Foundation National Center for Atmospheric Research, Earth Observing Lab, Boulder 80307, USA. ²Ann and H.J. Smead Aerospace Engineering Sciences, University of Colorado at Boulder, Boulder 80303, USA. ✉email: mhayman@ucar.edu

$$\alpha_i = \sum_{n=1}^N \int_{t_i}^{t_i+\Delta t} \rho_n(\tau) d\tau \quad (1)$$

where $\rho_n(t)$ is the photon flux at time-of-flight t for laser shot n , the bin number is defined by the interval $[t_i, t_i + \Delta t)$ and Δt is the histogram bin width in time-of-flight. A total of N laser shots are acquired per histogram. In this work we let t be time-of-flight relative to the most recent emitted laser pulse.

The standard approach to estimating backscatter amplitude is to place photon arrivals in a histogram (either immediately or during post-processing) and the observed photon counts in a bin is a Poisson random variable

$$y_i \sim \text{Poisson}(\alpha_i). \quad (2)$$

Given these photon count observations, the ubiquitous standard estimator for the backscatter amplitude is

$$\tilde{\alpha}_i \approx y_i \quad (3)$$

or if the photon flux is desired

$$\tilde{\rho}_i \approx y_i / (N \Delta t). \quad (4)$$

Because the photon counts in the histogram bin are Poisson distributed, the standard deviation in the observed counts is $\sigma_{y_i} = \sqrt{\alpha_i}$. If the signal-to-noise (SNR) is defined as the signal mean divided by standard deviation, then for a given bin

$$\text{SNR}_i = \sqrt{\alpha_i} = \sqrt{N \Delta t \bar{\rho}_i} \quad (5)$$

where $\bar{\rho}_i$ is the mean photon flux over the acquisition interval of the histogram defined

$$\bar{\rho}_i = \frac{1}{N \Delta t} \sum_{n=1}^N \int_{t_i}^{t_i+\Delta t} \rho_n(\tau) d\tau = \frac{\alpha_i}{N \Delta t}. \quad (6)$$

Operating under the assumption that the photon flux remains constant in the region of observation (although this assumption may not strictly hold for many operational remote sensing observations), it becomes evident that the precision of the standard estimator is directly linked to the histogram bin size. This connection demonstrates the problem of applying the standard estimator to high-resolution photon count data, where the reduction in SNR resulting from the use of small histogram bins and short integration times can only be mitigated by increasing the photon flux. Conventionally achieving higher photon flux entails employing higher laser power and larger collection apertures. However, this will tend to increase instrument cost, size, weight, and power (SWaP) requirements, make systems more complex. Additionally, it can push photon fluxes into nonlinear detection regimes, potentially compromising data accuracy and reliability.

It is important to recognize that as we push for higher resolutions, the precision of the standard estimator diminishes. However, it is equally important to recognize that the high-resolution data has no less information than that at coarser resolutions; in fact, the converse is true. When we bin data into coarser resolution information is lost. Moreover, the information content of high-resolution data is increasingly contained in the spatial clustering of those counts instead of the histogram count values. This limitation of the standard estimation approach becomes especially critical when dealing with sparse data, where the ability to leverage clustering information is essential for accurately estimating the backscatter signal.

For the purposes of this work, a sparse scene is characterized by instances in which $\alpha_i < 1$, or equivalently $\bar{\rho}_i < 1 / (N \Delta t)$, such that the probability of zero photon counts is relatively high and the probability of more than one photon count is relatively low. Thus sparse data can be the product of very high-resolution data acquisition or very low photon flux, but the distinction is not critical for the applications of this work.

Time-tagged data acquisition, sometimes referred to as Time Correlated Single Photon Counting (TCSPC), is a high resolution approach to acquiring and storing photon counting detector output¹⁰. In this mode of acquisition, photon counts registered from a detector are stored as time stamps, recording time-of-flight relative to the time of the transmitted laser pulse. This information is typically stored on a shot-by-shot basis. The technique was originally developed for measuring fluorescence decay functions¹¹. However it has since expanded into a number of applications where high timing resolution is required. Time-tagging is frequently employed in hard target lidar applications^{12–14}. The acquisition range resolution is determined by the precision of the onboard clock measuring the time-of-flight, enabling the captured range resolutions to be extremely high (down to centimeters or millimeters). In principle, when range capture resolution is infinitely fine, this approach avoids any information loss during the acquisition step. This ideal scenario represents an extreme case where *all* information content is carried in the clustering of the observed photon counts, with no information carried in the count quantity itself. While this time-tagged acquisition avoids any loss of information, we should acknowledge that it also introduces practical challenges when implemented in regular operations. One of the main challenges is the substantial data volumes generated by high-resolution data capture. This often requires large data storage to archive data and high data transfer rates to move the raw data off the lidar device or platform. This becomes particularly significant for instruments that run continuously for long time periods.

Seminal work in using time tag data for atmospheric lidar, a common application of distributed target lidar, was demonstrated in¹⁵ and more recently¹⁶. These instruments have captured atmospheric backscatter data at unprecedented resolutions. Consequently, they have the potential to provide important insights into the highly

dynamic and heterogeneous nature of many atmospheric targets. The influence of heterogeneous and dynamic targets on lidar data product accuracy have received relatively little attention, but there is some growing evidence that these factors can adversely affect remote sensing data. For instance, one study has shown that averaging over heterogeneous cloud structures can bias CALIOP backscatter and depolarization data products¹⁷. Additionally, heterogeneous cloud structure has been found to produce errors in estimates of cloud microphysical properties from passive satellite-based retrievals¹⁸. The ability to acquire high-resolution lidar data is an important step toward gaining a better understanding of these often ignored errors in real atmospheric scenarios.

Exploring high-resolution interrogation of the atmosphere is an important and largely unexplored scientific domain. However, it is important to note that the current products of the aforementioned instruments still represent an incomplete solution to investigate this domain. Although these works demonstrate high resolution *capture* of photon count data, a key distinction lies between the resolution at which data is captured and the resolutions actually supported by the information content. In essence, these instruments may simply be oversampling a noisy signal. While the time-tagging approach seems to offer unprecedented resolution in the capture of atmospheric structure, the remaining challenge lies in actually recovering lidar signal estimates while still leveraging the benefits of the acquisition mode. In both works mentioned, the high-resolution time-tagged data is eventually placed in histograms, and the standard estimation approach is employed to suppress shot noise and recover an estimated signal with acceptable precision. However, this requires significant reductions in resolution, and therefore largely negates the benefit of the high-resolution acquisition approach employed by these instruments.

Quantum Parametric Mode Sorting (QPMS) is a lidar technique often employing time tagging. It leverages mode matching criteria in nonlinear photonic crystals to improve the SNR in the optical signals reaching the detectors, as outlined in^{19,20}. While this technique may enable much better noise rejection, it comes at the cost of significantly lower total photon flux on the detectors due to the tightly constrained mode selection criteria. As a result, lidar profiles generated using QPMS in time-varying scenes will almost inevitably be sparse. It's worth noting that published works on this subject tend to refrain from providing details on the integration times compared to standard approaches and appear to largely focus on very close-range hard-targets characterized by high backscatter. This forces us to speculate regarding the possible fluxes expected in distributed target lidar. While most experiments with QPMS have focused on laboratory demonstrations with static targets, the possible and proposed applications for the technology have included environmental sensing problems from moving remote platforms where distributed target lidar would necessarily be employed (e.g. see²¹).

When developing sensors and data processing techniques for environmental sensing it's crucial to acknowledge that the scenes under interrogation are rarely static. This challenge becomes further compounded when the sensor is on a moving platform. It is important to note that laboratory demonstrations are able to interrogate static scenes over time scales that do not accurately reflect many of the proposed real-world applications. And in the instance of¹⁶ high *range* resolution atmospheric backscatter estimates were achieved, but at the apparent expense of time resolution. The highly dynamic nature of the clouds under investigation means that integration in time will smear out the image in range and largely negate the benefit of the sub-meter range resolution in the acquisition system. To remain fundamentally true to the notion of high-resolution distributed target lidar, it is essential to perform an accurate retrieval from sparse photon counting data that accounts for variations in both time and range. In order to maximize signal resolution, more advanced signal processing techniques are needed to recover useful signals from sparse photon counting data *and* non-stationary scenes without dramatically sacrificing the signal resolution in either axis. This is where Maximum Likelihood Estimation (MLE), and adaptations of it, can be employed to improve lidar data products by attempting to more efficiently leverage the information content, constraints, and uncertainties in the lidar signals.

The benefit of adopting an MLE based approach is that it potentially enables signal recovery by leveraging both the photon counts and their clustering to generate a signal estimate. MLE can thus be applied to time-tagged data directly, or similarly, to sparse and noisy histograms. This approach represents a realistic means of achieving better observations, rather than degraded ones, as we capture, and subsequently process, data at higher resolutions. While the raw observations appear noisy at high resolutions, the recovered signal estimates can still have low noise contamination, and more accurately reproduce the target structure compared to lower-resolution histogram estimator methods, which prioritize noise reduction at the expense of capturing accurate structure details.

The concept of applying MLE based techniques to process images dominated by Poisson noise is not new. Solutions have been, and continue to be developed, for a variety of applications including medicine^{22–24}, communications²⁵, astronomy²⁶ and ranging²⁷. Prior works on sparse photon counting problems in lidar have almost exclusively focused on single point target (hard target) ranging^{28,29}. However, this does not directly translate to challenges posed by volumetric (distributed) target scenarios, where the backscatter amplitude serves as the parameter, or intermediate parameter, of interest and is nonzero throughout the range profile. Such scenarios include the sensing of the atmosphere, snow characterization, and more. The distinctive contribution of this work lies in addressing the specific challenges posed in distributed target lidar systems and therefore fulfills the signal processing needs of the environmental sensing techniques described above.

Poisson Total Variation (PTV) is a penalized likelihood estimator for processing distributed target lidar data. The technique is inspired by signal processing approaches originally developed for medical imaging^{23,24} and has been demonstrated for atmospheric lidar in prior works^{8,9,30} where the focus is on retrieving data products at resolutions typical to atmospheric lidar, and photon counts are not sparse. In those studies, there was no concerted effort to evaluate the impact of processing raw data at higher resolutions than were already typical of the standard processing approach. Sparse data scenarios were intentionally avoided.

In this research endeavor, we show how a modified version of PTV can effectively process sparse data characterized by high temporal and range resolution. In this case we obtain backscatter estimates from the MicroPulse DIAL at resolutions as fine as $0.02 \text{ s} \times 0.75 \text{ m}$. This raw lidar data includes but is not limited to time-tagged data. Through simulations and analysis of actual atmospheric lidar data, we demonstrate that PTV is able to provide

more accurate estimates of photon arrival rate at a higher resolution than the standard histogram approach. This demonstrates its utility for processing time-tag and other high-resolution sparse photon counting data. By achieving this, PTV can serve as a useful tool for studying the atmosphere's heterogeneous and dynamic structure and its resulting impact on lidar data product accuracy. Importantly, our work completes the necessary components to leverage hardware described in^{15,16}. Our results indicate that sparse retrievals result in more accurate backscatter estimates than those achieved using PTV at lower resolutions, suggesting that nearly all atmospheric lidar data is over-averaged. However, the full impact of this over-averaging is still not known, highlighting the need for further investigations using this signal processing tool.

Poisson time tag noise model

Photon arrival times are described by a Poisson point process parameterized by the photon arrival rate $\rho(s)$ where s is the continuous time dimension (without reference to a laser shot event). As such, a sequence of photon detection times $\{s_i\}_{i=1}^I$ (assuming an ideal detector), where a total of I photon counts are detected, will be described by the probability distribution function (PDF) for a heterogeneous Poisson point process³¹

$$P(\{s_i = s_i\}_{i=1}^I) = \exp \left[- \int_0^{s_I} \rho(\tau) d\tau \right] \prod_{i=1}^I \rho(s_i). \quad (7)$$

In the equation above, we describe the detection process as a 1D time series, where in a lidar, there may be periodic laser firing. As such, there may be a desire to leverage correlation in the target structure across multiple laser shots. We can accordingly rewrite equation (7)

$$P(\{s_i = s_i\}_{i=1}^I) = \exp \left[- \sum_{n=1}^N \int_{x_n}^{x_{n+1}} \rho(\tau) d\tau + \int_{s_I}^{x_{N+1}} \rho(\tau) d\tau \right] \prod_{i=1}^I \rho(s_i), \quad (8)$$

where x_n refers to the absolute time of the n th laser shot and we let $x_1 = 0$ for simplicity. If the data is acquired over many laser shots, we assume that the second integral term in the exponential (a residual accounting for the time between the last observed photon and the end of the image) is negligible. Using this approximation, we can rewrite equation (8) as a 2D image estimation problem where we replace the absolute time variables s_i with time stamps relative to the most recent laser shot $t_i = s_i - x_n$ (for $x_n \leq t_i < x_{n+1}$) resulting in

$$P \left(\left\{ \{T_i = t_i\}_{i=1}^{I_n} \right\}_{n=1}^N \right) = \prod_{n=1}^N \left\{ \exp \left[- \int_0^{\Delta t_{RR}} \rho_n(\tau) d\tau \right] \prod_{i=1}^{I_n} \rho_n(t_i) \right\}. \quad (9)$$

where Δt_{RR} is the repetition period of the laser ($\Delta t_{RR} = x_{n+1} - x_n$) and $\rho_n(t)$ is now a 2D variable with continuous dimension in time-of-flight t and discrete dimension in laser shot index n .

To employ MLE, we seek to find an estimate $\tilde{\rho}_n(t)$ that minimizes the negative log-likelihood (NLL) of the detection noise model. This is the optimization loss function that effectively evaluates how well the estimate $\tilde{\rho}_n(t)$ fits the observed time tag data

$$\mathcal{L}_{TT} \left(\tilde{\rho}_n(t); \left\{ \{T_i = t_i\}_{i=1}^{I_n} \right\}_{n=1}^N \right) = \sum_{n=1}^N \left(\int_0^{\Delta t_{RR}} \tilde{\rho}_n(\tau) d\tau - \sum_{i=1}^{I_n} \ln \tilde{\rho}_n(t_i) \right). \quad (10)$$

Equation (10) represents the generalized fit loss for estimating a photon arrival rate $\tilde{\rho}_n(t)$ from observations of time tag data. This provides a direct mechanism for evaluating the accuracy of candidate solutions for $\tilde{\rho}_n(t)$. Any variety of basis functions may be used to perform this estimate. For example, polynomial and spline functions would be reasonable options depending on the specific description of the problem of interest. Later in this work, we will employ PTV, where the basis functions are a set of piecewise constant functions. This is also a natural description of the standard histogram approach where each bin is treated as a constant over the bin width. Piecewise constant functions will additionally enable further simplification of the NLL.

As is the case in prior works^{8,30}, $\tilde{\rho}_n(t)$ may also be a forward modeled function of other retrieved variables in order to obtain denoised estimates of the derived data products rather than the photon arrival rate itself. In this work, we focus on recovering the photon arrival rate because it is an essential component of any quantitative estimate (the retrieved quantity that maps directly onto observations), and therefore many of the concepts covered here broadly apply to a number of lidar architectures and applications.

Poisson total variation

Poisson Total Variation is a signal processing approach that draws inspiration from medical imaging techniques^{23,24}. The technique was first applied to lidar data for High Spectral Resolution Lidar (HSRL) retrievals of backscatter and extinction coefficients³⁰ and later adapted to obtain denoised retrievals of water vapor profiles from MicroPulse DIAL (MPD) data⁸. Under the PTV approach, variables are represented as images consisting of discrete pixels, which we assume can be reasonably approximated using a basis of piece-wise constant functions. This solution basis is imposed by applying total variation regularization to the estimated lidar signal, which is a penalty for each change in value between pixels (i.e. increase in complexity) in the estimated image. In concept, this forces the optimizer to prefer simpler images, where responding to a change in signal

must produce a sufficient gain in fit quality to overcome the penalty of the change. Under this framework, the minimized objective function becomes a sum of the noise model and the total variation regularization

$$\mathcal{O}(\mathbf{x}; \mathbf{y}) = \mathcal{L}[f(\mathbf{x}); \mathbf{y}] + \eta \|\mathbf{x}\|_{TV} \quad (11)$$

where $f(\mathbf{x})$ is the forward model that maps the estimated variable image \mathbf{x} to a photon flux image $\tilde{\rho}$, \mathbf{y} are the 2D photon count observations, η is a scalar tuning parameter that sets the amount of total variation penalty for the 2D image \mathbf{x} consisting of I rows and J columns, and $\|\cdot\|_{TV}$ is the l_1 -based, anisotropic total variation defined³²

$$\|\mathbf{x}\|_{TV} = \sum_{i=1}^{I-1} \sum_{j=1}^{J-1} (|x_{i+1,j} - x_{i,j}| + |x_{i,j+1} - x_{i,j}|) + \sum_{i=1}^{I-1} |x_{i+1,J} - x_{i,J}| + \sum_{j=1}^{J-1} |x_{I,j+1} - x_{I,j}|. \quad (12)$$

In order to minimize the objective function, we employ a version of Sparse Poisson Intensity Reconstruction Algorithm (SPIRAL), SPIRAL-TV^{23,24}, in combination with Fast Iterative Shrinkage/Threshold Algorithm (FISTA) (specifically fast gradient projection-FGP-in³²). Each optimization step is calculated

$$\mathbf{x}^{k+1} = FGP \left\{ \mathbf{x}^k - \frac{1}{\tau_k} \nabla \mathcal{L}[f(\mathbf{x}^k); \mathbf{y}], \frac{\eta}{\tau_k} \right\} \quad (13)$$

where τ_k is effectively the gradient descent step size and determined using SPIRAL-TV. The acceptance criteria of each step is similarly determined using SPIRAL-TV.

The determination of the optimal total variation penalty η is achieved using holdout cross validation which is described in more detail in the subsection below.

When we employ PTV, the estimated variable is approximated as a set of piecewise constant functions and requires an assumed discrete grid for the solution basis. As a result, the solution is defined for discrete intervals of index p in time-of-flight, and q in laser shots, such that $\tilde{\rho}_n(t_i) = \tilde{\rho}_{p,q}$ when $t_p \leq t_i < t_{p+1}$ and $q \leq n < q+1$. Given the discrete solution, the time tag loss function in equation (10) simplifies for PTV to

$$\mathcal{L}_{PTV}(\tilde{\rho}; \mathbf{y}, N) = \sum_{q=1}^Q \sum_{p=1}^P [N_q \tilde{\rho}_{p,q} \Delta t_p - y_{p,q} \ln \tilde{\rho}_{p,q}]. \quad (14)$$

where there are P time-of-flight bins in the estimated backscatter, with bin width Δt_p , Q total laser shot bins, and $y_{p,q}$ are the number of photon counts that fell into the (p, q) grid point as defined by its arrival time t_p and laser shot n .

Note that the loss function in equation (14) is effectively the loss function employed for a Poisson random number where \mathbf{y} is a histogram of photon counts. Because of the discretization in the PTV solution basis, there is no practical reason to retain the time tag data in its raw form. In applying PTV, raw time tag data is naturally arranged into histograms defined by the resolution of the retrieved image $\tilde{\rho}$. However, this PTV histogram resolution can be *much* finer than would be typical if we aimed to estimate $\tilde{\rho}$ directly from histogram bin counts. This very fine histogram resolution is thus also flexible in responding to the observed scene.

For general instances of time tag estimation using continuous parametric fitting (such as polynomials or splines), the time tag loss function in equation (10) can be employed to avoid loss of information caused by binning the photon counts in a histogram. However, because PTV requires that we select a solution grid size to define the basis image, under the assumption that $\tilde{\rho}$ is constant within a grid point, there is no practical advantage to processing the raw time tags. Instead, we will use equation (14) when performing PTV retrievals in the next sections.

We will show, however, that data can be recovered at much higher resolutions and more accurately using PTV than when employing standard histogram methods. The best retrievals from PTV will be obtained from very sparse histogram data that otherwise has no immediately discernible signal using the standard approach at the same resolution.

Holdout cross validation

The total variation penalty η is effectively a constraint on the amount of complexity allowed in the estimate of \mathbf{x} . Small values of η allow for more variation in the estimate, and can thus respond to rapid changes in the state variable but are also more likely to fit noise. By contrast large values will tend to suppress variations in the estimate. The optimal value for η will vary based on signal-to-noise and the variability of the scene under investigation. As a result, in the PTV framework, η is optimized on a scene-by-scene basis. In order to achieve this optimization, we employ holdout cross validation, where the solution obtained for a particular value of η is compared to independent identically distributed (i.i.d.) holdout data. The evaluation of η is conducted using the same loss function as during the fitting process $\mathcal{L}(\tilde{\rho}; \mathbf{y}_v, N_v)$ where now the observed data is the held-out validation data. Lower values of the loss function indicate a better fit and more optimal total variation penalty. In this work we obtain the optimal value of η by evaluating a series of values and selecting the value that gives the smallest validation loss (if the minimum is an end point, we extend the search space).

For photon counting, there are two methods for obtaining the holdout cross-validation data \mathbf{y}_v . In each case, data from the total observed signal must be split into separate fit (used for estimation) and validation (used for

optimizing η) datasets to ensure independence. In the case of Poisson distributed observations, fit and validation data can be split directly from the original dataset using Poisson thinning²⁴. In the case of histograms, Poisson random data is split using a binomial random number generator as described in³³. Thinning time tag data resulting from a Poisson point process is also possible using a Bernoulli random number generator for each time tag. Each time tag is placed into a set according to the outcome of its corresponding Bernoulli trial (e.g. 0—fit data, 1—validation). We should note that the process of thinning will reduce the amount of signal available for direct processing, but we have found that signal processing that is analytically optimized for the scene seems to always produce better results than the alternative.

In this work, we employ “manual” thinning to generate a fit and validation datasets. In this approach, data from alternating laser shots is stored in separate fit and validation datasets. This is a reasonable approach to achieving i.i.d. fit and validation data because the laser sample rate (in this case 8 kHz) is much higher than the resolvable variations in the observed scene (in this case on the order of 10–100 s of Hz). However, this highlights the potential drawback of manual thinning if the scene is not oversampled by the laser, i.e. from fast-moving platforms, in highly dynamic scenes or lower repetition rate lasers. The benefit of the manual thinning approach is that it allows us to apply holdout cross-validation even when the noise model is not Poisson, which is potentially an important eventual application for the methods described here. For instances of analog detection or where detector nonlinear effects result in non-Poisson photon counts, manual thinning still presents a viable option for holdout cross-validation.

Coarse-to-fine

Directly processing fine-resolution (sparse) photon count data with PTV does not tend to produce a reliable solution. In³⁰ it is noted that zero-value histogram bins result in non-unique solutions in PTV retrievals. In previous work, we leveraged a technique termed coarse-to-fine (PTV-CF) in order to obtain more accurate water vapor retrievals over large time intervals⁸. Here we show that a similar approach can be leveraged in order to obtain high-resolution estimates of backscatter structure even when the photon counts are small (typically one) and sparse (most bins are zero) at the retrieved resolution.

We term this high-resolution coarse-to-fine version of PTV as PTV-CF-HR. Simply stated, this processing approach starts with coarse images that have been binned at a low enough resolution to have non-zero photons in all pixels. That first estimate is used as the initial condition for the next estimate at a finer resolution, and the process is repeated. This approach is described by the flow diagram in Fig. 1. In the figure, key variables are set in color and the arrows define the flow of those variables into each processing step. Blue corresponds to the backscatter photon flux estimate, yellow is the resolution and the photon counts are red. We start with a set of photon count time tags and an inverse ordered set of integer multipliers that define the coarse-to-fine resolution steps relative to the base resolution of the final image. The photon counts are manually thinned by laser shots and binned into a separate fit and validation data set at the finest (final) processing resolution (In the case of the MPD instrument, we allow this to be defined by the rate of the optoelectric switch used to change between transmitting online and offline laser wavelengths).

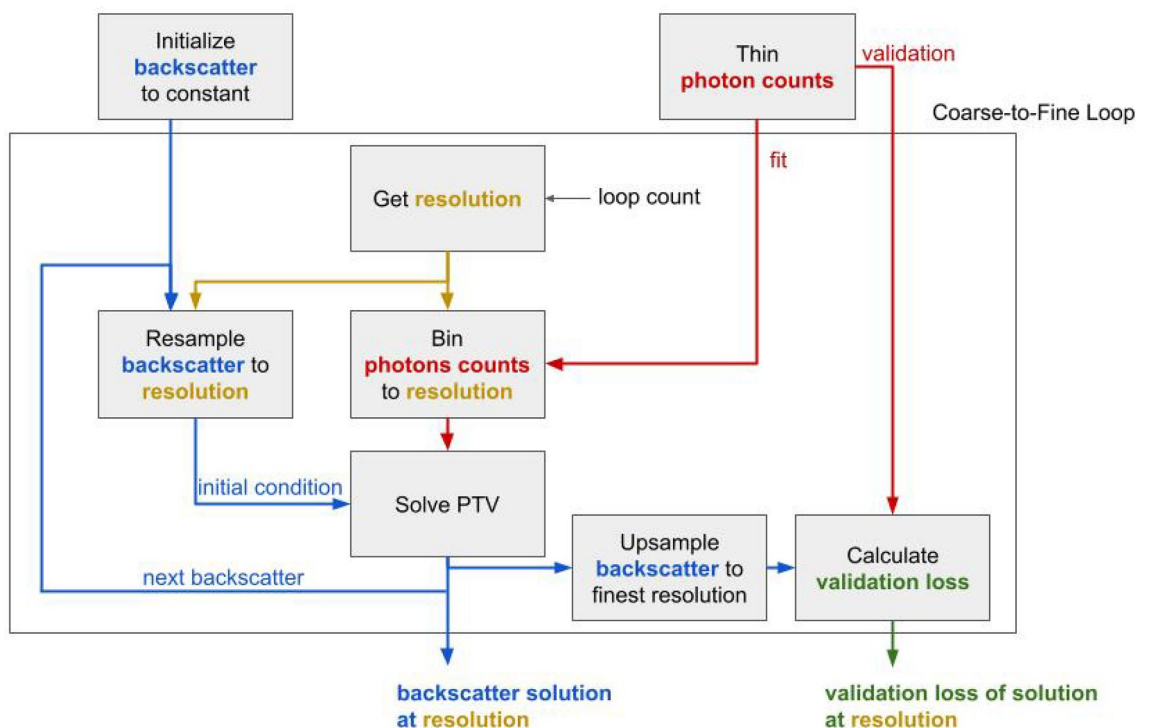


Figure 1. Flow diagram of the high-resolution PTV-CF-HR algorithm.

The initial estimate of the backscattered photon flux is set to a constant (generally 1 kHz). For each successive finer resolution, the fit photon counts are binned from the raw time tags to the new finer resolution. In addition, the initial condition for PTV is the previous solution resampled to the new resolution. The solution obtained from PTV is evaluated against the validation photon count data by first upsampling the solution to the finest (final) resolution and comparing it (using the NLL) to the validation data similarly binned at the finest base resolution. This final estimate is a refined version of the initial estimate where significant high-resolution features are allowed to persist while noise is largely ignored via total variation regularization.

In⁸, coarse-to-fine was performed incrementally only on the water vapor field. The water vapor estimate was resampled to project onto a base resolution of 1 min by 37.5 m data. In this work, we adjust the resolution of the photon count histogram and the retrieved photon arrival rate $\tilde{\rho}$. In addition, instead of incrementally decreasing bin size (which would require significantly more coarse-to-fine steps), in this work, we typically decrease bin size by a factor of two at each step.

Simulation

In this section, we provide a demonstration of PTV-CF-HR on a simple simulation of randomly generated rectangular patches (where there is overlap, their arrival rates are summed). This enables us to demonstrate the concept and establish the use of the validation NLL as a metric for evaluating the method on actual atmospheric data in the next section. This is important because on actual observations no truth is available for computing root-mean-square error (RMSE). The simulated scene is sampled using a 10 kHz laser repetition rate. Photon arrivals are simulated as time tagged but then binned to a base resolution of 1 ns (approximately 0.15 m) for PTV processing. In order to create a validation data set, photon counts from every other laser shot is held out for cross-validation and optimization of the total variation regularizer η . An example of the simulated true photon flux is shown in the top left panel of Fig. 2. The fit photon count data is processed at a variety of resolutions which are integer scales of the base resolution (applied to both range and time). The fit photon count data is placed in a histogram at the same scaled resolution.

To perform the PTV-CF-HR estimate, we employ the coarse-to-fine solution described in Fig. 1. In this approach, the data is first processed at low resolution using PTV and the solution from that processing step is the initial condition for the next higher-resolution step. This helps ensure that the solution converges in cases of sparse photon count observations where the optimization problem is not strictly convex.

Figure 2 shows the simulated true photon flux along with estimates of the photon flux using a variety of histogram resolutions (reported as multipliers relative to the base resolution) and the result of PTV-CF-HR at 10x the base resolution. As expected, estimates of photon flux are quite noisy when histogram bins are small. As we increase the bin sizes, the noise is suppressed and the photon flux estimates improve in the histogram approach. However, at larger bin sizes, the ability to resolve sharp edges in the image is diminished, resulting in the smearing of the actual structure. As a result, errors begin to increase beyond a bin resolution of 200x. This may be understood as a result of features in the actual image poorly aligning with the predefined histogram grid, or having structure that is finer than the grid. By contrast, PTV-CF-HR, is largely able to resolve the fine-scale features and edges at 10x resolution, while simultaneously suppressing random noise throughout the image.

We analyze the retrieved estimates of the photon flux using both PTV-CF-HR and the standard histogram approach by calculating the RMSE—which is only possible because we have access to absolute truth in simulation—and the validation negative log-likelihood—which can be applied in actual atmospheric measurement cases as long as we have hold out data. These error metrics are shown in Fig. 3 for both estimation approaches at a variety of resolutions (shown as scale factors relative to the base resolution). In the case of the validation NLL, the actual values are adjusted so the minimum point (which is negative) is at 1, allowing us to display the result on a log axis.

Note that both the RMSE (top panel) and validation NLL (bottom panel) show similar trends. This demonstrates how validation NLL can be a useful analytical proxy for error in lidar retrievals where we usually lack a truth reference. We should note some caveats, however. The validation NLL is only accurate if the noise model used to calculate the NLL is an accurate representation of the detection statistics, and the noise is uncorrelated between fit and validation observations. Also, this error is only reflective of how well the retrieval estimate represents the observations. It cannot account for errors in calibration or inaccuracies in the assumed physical model of the instrument.

The PTV-CF-HR and histogram approach to estimating photon flux have nearly identical errors at low resolutions (above a scale factor of 400). For the histogram approach, the lowest RMSE and validation NLL occur at a scale factor of 200 where the two processing approaches have begun to diverge. At an even finer resolution, PTV-CF-HR continues to become more accurate, due to its improved ability to resolve sharp features in the image, while the histogram approach becomes noisier at finer resolution, and due to this noise, errors begin to increase. We also note that below a scale factor of 20, there is little additional improvement in the PTV-CF-HR retrieved signal. This appears to be the performance limit of the retrieval for the given observations.

This analysis shows that by applying a penalized likelihood estimator like PTV, we can achieve lower error at a higher resolution than the standard histogram processing approach.

Atmospheric data

To further demonstrate PTV-CF-HR, we applied it to atmospheric lidar data and used the validation NLL for evaluation. The NSF-NCAR MicroPulse DIAL (MPD)³⁴ was operated with a custom time-correlated single photon counting (TCSPC) module with 5 ns time-of-flight resolution. The laser repetition rate is 8 kHz (3 μ J per pulse) with the DIAL system switching between online and offline water vapor channels at a period of 50

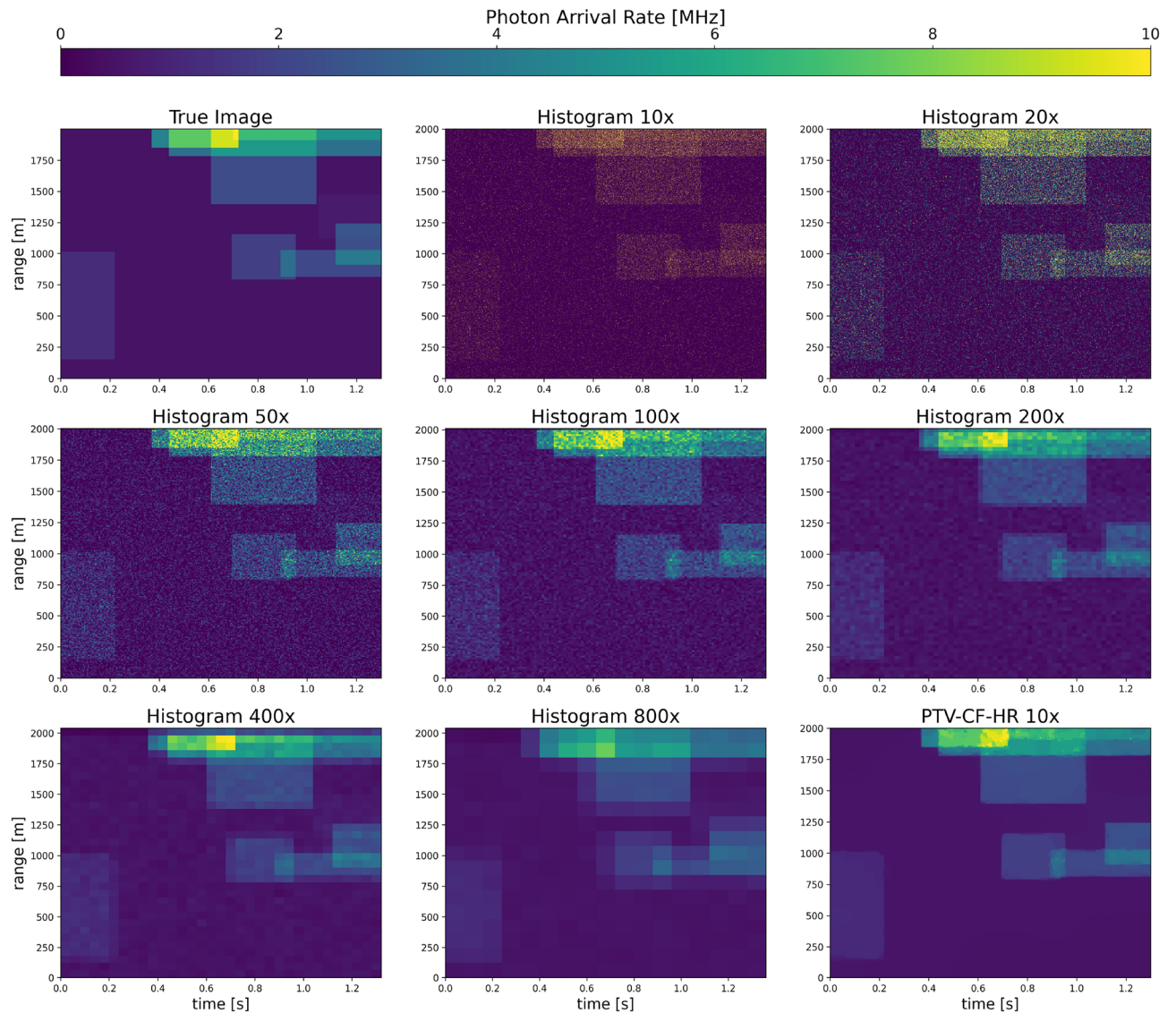


Figure 2. True image (top left) compared to histogram estimates of photon flux for incrementally increasing bin size (indicated as multipliers of the base 1 ns range and 0.1 ms time resolution). At fine resolution, the histograms are noisy while at coarse resolution, images smear out. The lowest error of the histogram approach is at 200x bin size. The bottom right image shows the result obtained by PTV-CF-HR at a resolution of 10x.

Hz. As a result of this hardware design, to avoid blank laser shots, we chose to bin the data for processing to the finest base resolution of 20 ms (80 shots per wavelength channel) in time and 5 ns (0.75 m) in time-of-flight.

Here we examine data from August 19, 2021³⁵ over a small altitude and time period (approximately 1.9–3.3 km for 5 minutes). During this time we observe falling ice, the melting layer, and subsequent falling rain, which presents an interesting backscatter target for analysis. All of the photon arrival rates from this scene appear to be well below the nonlinear regime of the single photon avalanche photodiode detector (with a dead time of approximately 25 ns).

In order to estimate the atmospheric backscatter signal, we employ a forward model that encapsulates the background counts and laser pulse width. Thus the detected photon arrival rate is

$$\tilde{\rho}_n(t) = L(t) * \hat{\rho}_n(t) + \hat{b}_n \quad (15)$$

where $L(t)$ is the square laser pulse (0.625 μ s or 93.75 m) and \hat{b}_n is the background counts estimated using PTV-CF-HR on high altitude photon counts near 18 km (which are assumed to contain negligible atmospheric backscatter signal). In the results we consider here, the solutions $\hat{\rho}$ are shown in the images, but the validation NLL is still evaluated based on the forward model $\tilde{\rho}$. In this way, the relatively long square laser pulse can be deconvolved from the atmospheric signal to obtain structure at a much finer resolution.

In performing this analysis, we consider a few different coarse-to-fine paths, where data is initially binned at different resolution ratios in time and range relative to the base resolution dictated by time of flight resolution (5 ns) and online-offline sampling period (20 ms). At each up-sampling step, the pixel size was reduced in

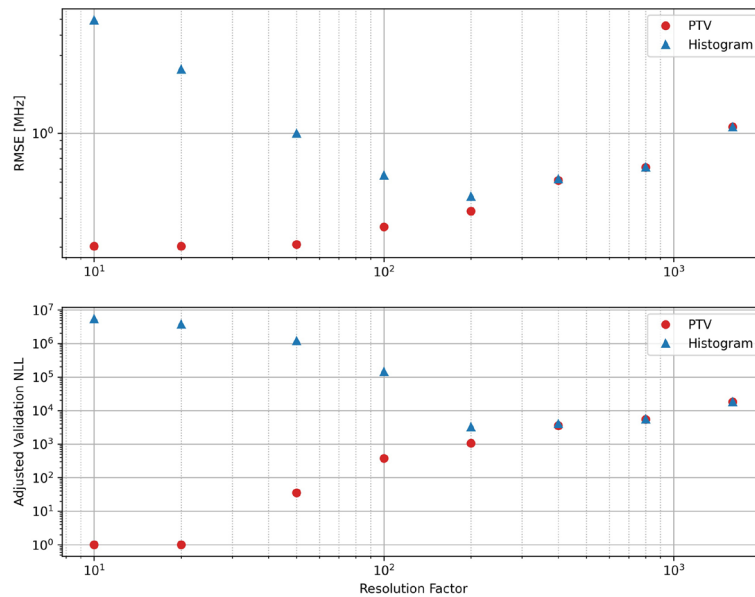


Figure 3. Evaluation of PTV processing approach compared to histogram method as a function of data resolution. RMSE was calculated using the true backscatter flux of the simulation (top). Adjusted validation NLL (bottom) is calculated relative to one plus the minimum NLL. Resolution is reported as multipliers of the base 1 ns range resolution and 0.1 ms time resolution (larger values indicate coarser resolution).

both dimensions by a factor of two, thus each resolution step is a power of two relative to the base resolution. The validation NLL of each up-sampled step is shown for each of these time:range processing cases as a function of time and range resolution in Fig. 4. The processing result that achieves the lowest validation NLL is 8:1 (bottom row, second from the right, marked with the white x). At present, we cannot definitively claim to know the reason for this (and therefore have prior knowledge of the best processing option), though it does appear to correspond to the case where the vertical and horizontal pixel dimensions are close to the same. We compare the final validation NLL of all these processing cases and include the optimal standard approach (Hist.) as shown in Fig. 5. In order to perform this comparison, we determined the optimal MCS bin resolution based on that which produces the lowest validation NLL. We note that all of the PTV-CF-HR cases produce a lower validation NLL than the optimal histogram approach.

The results of all of these backscatter estimates $\hat{\rho}$ are shown in Fig. 6 which includes the optimal histogram solution (10.24 s \times 48 m) and the raw photon time tags as well. Notably, the optimized histogram solution still retains noise in low signal regimes and significantly degrades the ability to resolve high-resolution structure in higher signal regimes. We know from the validation NLL in Fig. 5, that this solution has higher error than the

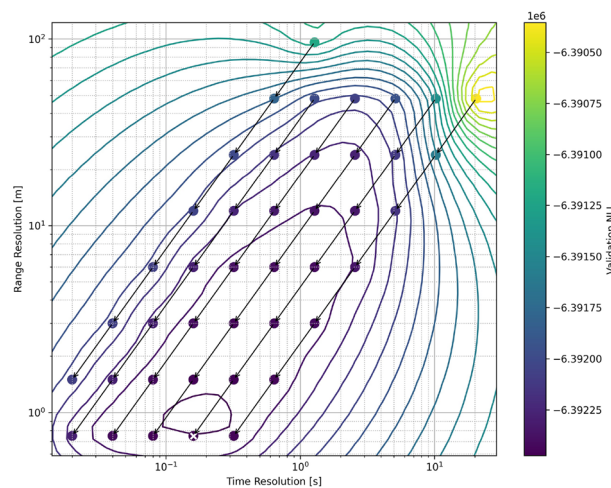


Figure 4. Map of validation negative log-likelihood for coarse-to-fine trajectories relating to different time:range bin ratios of the raw photon count data. The white x indicates the lowest validation NLL of the candidate solutions.

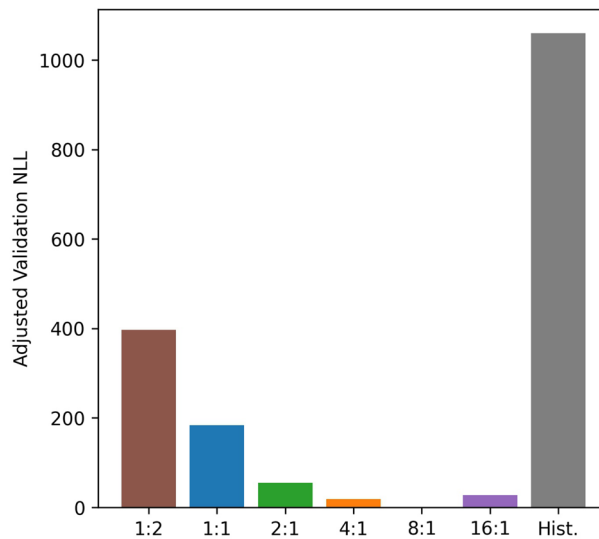


Figure 5. Comparison of final estimate validation log-likelihood based on different time:range bin ratios. Also compared is the optimal histogram binning solution representing the best-case standard approach.

PTV-CF-HR solutions, but Fig. 6 makes it visually clear that the standard processing approach results in a lower quality (higher noise, lower resolution) estimate of the lidar backscatter signal.

The average PTV processing times for each coarse-to-fine step in the above analysis are shown as a function of pixel count in Fig. 7. PTV is expected to take longer to achieve convergence when processing larger images, which is clearly reflected in the data. The coarsest resolutions (smallest images) complete in minutes while the finest resolutions (largest images) took nearly 12 h to complete. Note that the rate of increase starts to diminish just below 10^6 pixels. In general we would expect the processing time to continue to increase with pixel count, but because each coarse-to-fine instance is initialized by the result of the previous step, it is likely this roll-off is a product of the initial condition being near the final solution.

The processing times required for PTV-CF-HR are clearly much larger than that for the standard histogram method. However this presents somewhat of a false comparison, as the histogram method cannot produce signal reconstructions at resolution and accuracy comparable to PTV-CF-HR. Thus if high resolution products are the objective, increased computational cost is inevitable. While the processing times shown here are clearly not sufficient for real time processing, they still enable scientifically useful study of heterogeneous structure in atmospheric targets in a way that was not previously possible with standard processing approaches.

There may be opportunities to increase the processing speed of PTV-C2F-HR. Better tuning of stopping criteria in the Spiral-TV instance would ensure that the optimization routine does not run for extended periods of time while achieving small improvements in the estimate. Also, the PTV software used in this work relies on a CPU based implementation. More recent developments in GPU based PTV processing may produce a substantial increase in processing speed⁹.

While the 8:1 coarse-to-fine path produces the best final result, there is still potential to process the data at finer temporal resolution. To evaluate the impact of this, we performed coarse-to-fine in only the time dimension. Figure 8 shows this analysis applied to all of the cases where we already obtained a minimum in-range resolution (75 cm) as dictated by the acquisition clock. In this scene, the validation NLL continues to decrease at finer temporal resolution for all of the considered cases. This suggests that there may be benefits in processing the data at still finer than 20 ms resolution. However, the impact if this final coarse-to-fine processing is relatively small. The order of performance remains the same as with the original time:range bin resolutions. In most cases, this remains true when comparing the worst and best cases. For example, note that the 8:1 solution before processing at higher temporal resolutions still has a lower validation NLL than the next best solution (4:1) even at its highest temporal resolution. This suggests that the initial condition remains an important parameter in obtaining the best possible solution image. It further emphasizes that, while all of the PTV-CF-HR cases provide superior retrievals to the standard method, more work is needed to determine the optimal initial coarse-to-fine resolution.

It is important to recognize that throughout the analysis of the scene presented here, the validation NLL continues to decrease as the data product resolution becomes finer. As we move to progressively finer resolutions, we should eventually expect to see the validation NLL cease to improve (or perhaps even become slightly worse) when information captured in the observational data is no longer sufficient to provide higher-resolution observations. That this does not happen, suggests that the information content captured in this scene supports much higher resolutions than are typically associated with atmospheric lidar data. That is, in effect, standard lidar resolutions are over-averaging photon counting data. For the relatively low power MPD, typical time resolutions are 1 minute with a range resolution of 37.5 m. However, the power of the lidar (and the noise sensitivity of its final data product) will tend to scale the supported resolutions so that higher power systems operating at 1-second resolution, could still be significantly over-averaging as well. Ultimately, over-averaging is driven by two factors,

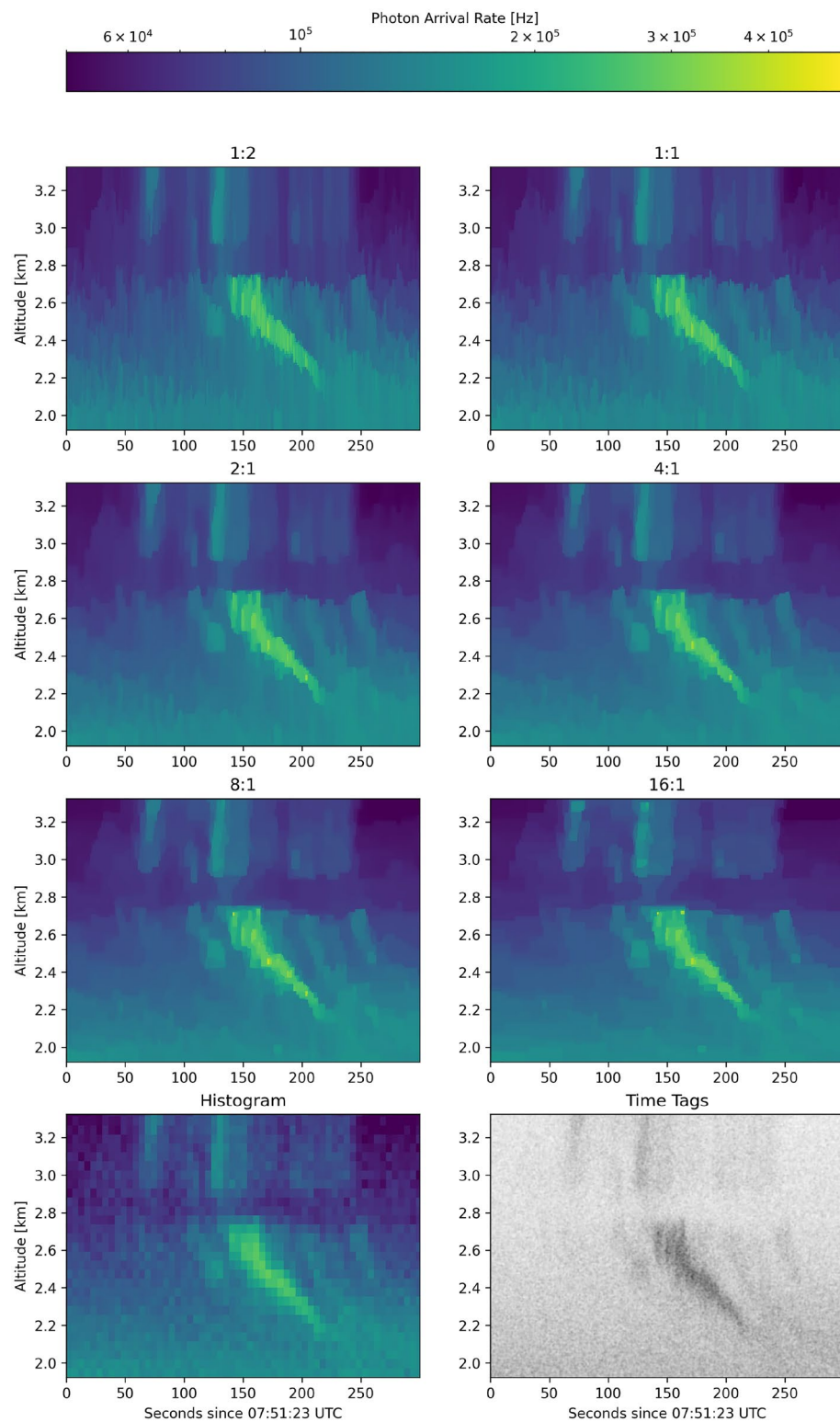


Figure 6. Comparison of PTV-CF-HR estimated photon arrival rate based on the relative raw data binning (time:range) compared to the base resolution of $0.02 \text{ s} \times 0.75 \text{ m}$. Based on validation NLL, the 8:1 case is the best result. Also shown is the optimal histogram solution representing the best-case standard approach and a scatter plot of the raw photon time tags.

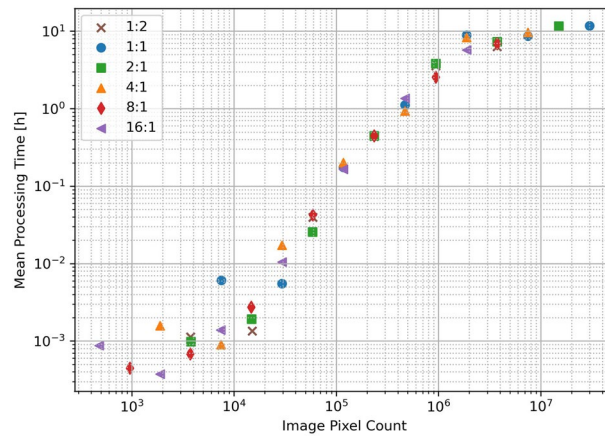


Figure 7. Average processing time of PTV routine as a function of pixels in the image.

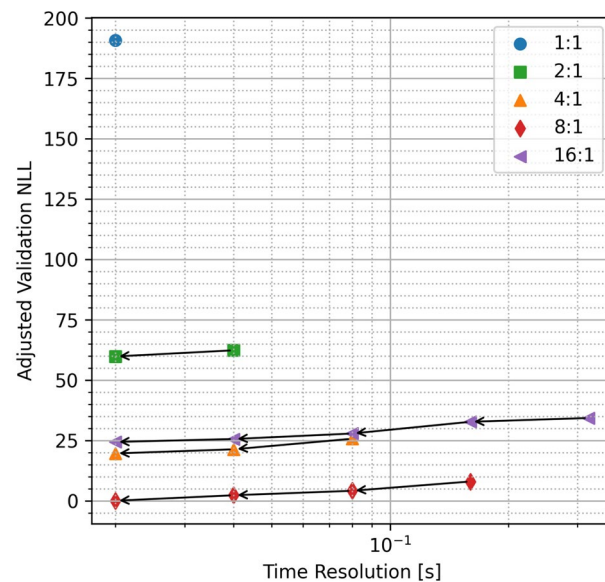


Figure 8. Comparison of finest temporal resolution estimates based on initial coarse-to-fine time:range bin ratios. Processing at finer temporal resolution further improves the validation NLL.

the spectrum of dynamic/heterogeneous structure in the scene under investigation (potentially extremely high for clouds) and the ability of the captured signals of that structure to rise above noise limits.

Conclusion

In this study, we have demonstrated a novel approach for applying maximum likelihood estimation to sparse, distributed target data obtained from photon-counting lidar systems. We have specifically focused on using a coarse-to-fine methodology coupled with Poisson Total Variation, termed PTV-CF-HR, to achieve high-resolution retrievals of atmospheric backscatter. This was accomplished using low-power lidar instrumentation with raw TCSPC data binned to a minimum temporal resolution of 0.02 s and a range resolution 0.75 m.

Our findings show that PTV-CF-HR outperformed the standard histogram approach to estimating lidar backscatter signals as shown in both the simulated and atmospheric processing. Notably, our results reveal that retrievals applied to higher-resolution data have improved validation performance, compared to coarser resolutions. This result suggests that atmospheric lidar data may generally be over-averaged at resolutions more typical of atmospheric lidar, although further investigations are needed to understand its impact on data product accuracy.

While the research presented here has primarily focused on retrievals of backscatter photon arrival rate, it can also be leveraged for atmospheric data products. This capability has been demonstrated with PTV at much lower resolutions where the data is not sparse^{8,30}. Combining the forward model retrieval approach with PTV-CF-HR could be useful for processing low signal-to-noise lidar observations; such as recovering the fine-scale structure of polar mesospheric clouds and sodium layers in the upper atmosphere. Additionally, it may prove

beneficial for retrievals leveraging Quantum Parametric Mode Sorting (QPMS) architectures in environmental sensing where captured signals are likely to be sparse over dynamically evolving scenes.

In addition, PTV-CF-HR serves as a tool for investigating lidar sensor performance and errors resulting from coarse-resolution acquisition and analysis on observational data collected from dynamic scenes. The effect of averaging lidar signals over heterogeneous cloud structures is rarely discussed in the literature, and there has been very limited investigation of its potential impact. Some analysis has suggested that space-based lidar products may be biased by heterogeneous scenes¹⁷ and there are similar indications for other remote sensors as well¹⁸, but there is currently very limited ability to investigate this effect directly. Combining PTV-CF-HR with high-resolution acquisition systems (as described in previous works^{15,16} and implemented on the MPD here) offers a pathway to gaining further understanding of remote sensor performance and accuracy.

While this work shows the potential for estimation techniques derived from Maximum Likelihood Estimation (MLE) applied to lidar data processing, we also highlight the need for further research. In particular, we focused on cases where non-ideal detector behavior can be reasonably neglected. Given that likelihood estimators rely on an accurate statistical model of the detection process, achieving accurate recovery of tropospheric cloud structure requires a noise model that can properly account for the non-ideal aspects of the detection processes. Further, the highest performance result obtained from PTV-CF-HR depended on the initial coarse-to-fine resolution selected. While it is possible to systematically evaluate a range of possible initial resolutions, this can be computationally expensive. Additional work that allows us to avoid this step would enhance the practicality of PTV-CF-HR.

Data availability

The data used to generate this paper is archived and publicly available³⁷. Atmospheric data is from raw MPD instrument which is publicly available³⁵. Analysis code used for this publication is available at <https://github.com/NCAR/Signal-Estimation-Sparse-Data>.

Received: 14 September 2023; Accepted: 23 April 2024

Published online: 06 May 2024

References

- Hadfield, R. H. Single-photon detectors for optical quantum information applications. *Nat. Photon* **3**, 696–705. <https://doi.org/10.1038/nphoton.2009.230> (2009).
- Zoller, P. *et al.* Quantum information processing and communication. *Eur. Phys. J. D* **36**, 203–228. <https://doi.org/10.1140/epjd/e2005-00251-1> (2005).
- Hsieh, S. S., Leng, S., Rajendran, K., Tao, S. & McCollough, C. H. Photon counting ct: Clinical applications and future developments. *IEEE Trans. Radiat. Plasma Med. Sci.* **5**, 441–452. <https://doi.org/10.1109/TRPMS.2020.3020212> (2021).
- Hadfield, R. H. *et al.* Single-photon detection for long-range imaging and sensing. *Optica* **10**, 1124–1141. <https://doi.org/10.1364/OPTICA.488853> (2023).
- Boksenberg, A. & Burgess, D. An image photon counting system for optical astronomy. *Adv. Electron. Electron Phys.* **33**, 835–849. [https://doi.org/10.1016/S0065-2539\(08\)60798-2](https://doi.org/10.1016/S0065-2539(08)60798-2) (2018).
- Rapp, J., Ma, Y., Dawson, R. M. A. & Goyal, V. K. Dead time compensation for high-flux ranging. *IEEE Trans. Signal Process.* **67**, 3471–3486. <https://doi.org/10.1109/TSP.2019.2914891> (2019).
- Rapp, J., Ma, Y., Dawson, R. M. A. & Goyal, V. K. High-flux single-photon lidar. *Optica* **8**, 30–39. <https://doi.org/10.1364/OPTICA.403190> (2021).
- Marais, W. J. & Hayman, M. Extending water vapor measurement capability of photon limited differential absorption lidars through simultaneous denoising and inversion. *Atmos. Meas. Tech.* **15**, 5159–5180. <https://doi.org/10.5194/amt-15-5159-2022> (2022).
- Hayman, M., Stillwell, R. A., Karboski, A., Marais, W. J. & Spuler, S. M. Global estimation of range resolved thermodynamic profiles from micropulse differential absorption lidar. *Opt. Express* **32**, 14442–14460. <https://doi.org/10.1364/OE.521178> (2024).
- Becker, W. *Advanced Time-Correlated Single Photon Counting Techniques* 1st edn. (Springer, 2005).
- O'Connor, D. V. & Phillips, D. *Time-correlated single photon counting* 1st edn. (Academic press, 1984).
- Massa, J. *et al.* Time-of-flight optical ranging system based on time-correlated single photon counting. *Appl. Opt.* **37**, 7298–7304. <https://doi.org/10.1364/AO.37.007298> (1998).
- McCarthy, A. *et al.* Kilometer-range, high resolution depth imaging via 1560 nm wavelength single-photon detection. *Opt. Express* **21**, 8904–8915. <https://doi.org/10.1364/OE.21.008904> (2013).
- Pawlikowska, A. M., Halimi, A., Lamb, R. A. & Buller, G. S. Single-photon three-dimensional imaging at up to 10 kilometers range. *Opt. Express* **25**, 11919–11931. <https://doi.org/10.1364/OE.25.011919> (2017).
- Barton-Grimley, R. A., Stillwell, R. A. & Thayer, J. P. High resolution photon time-tagging lidar for atmospheric point cloud generation. *Opt. Express* **26**, 26030–26044. <https://doi.org/10.1364/OE.26.026030> (2018).
- Yang, F. *et al.* A time-gated, time-correlated single-photon-counting lidar to observe atmospheric clouds at submeter resolution. *Remote Sens.* **15**, 1500. <https://doi.org/10.3390/rs15061500> (2023).
- Alkaseem, A. *et al.* Effects of cirrus heterogeneity on lidar caliop/calipso data. *J. Quant. Spectrosc. Radiat. Transf.* **202**, 38–49. <https://doi.org/10.1016/j.jqsrt.2017.07.005> (2017).
- Arola, A. *et al.* Aerosol effects on clouds are concealed by natural cloud heterogeneity and satellite retrieval errors. *Nat. Commun.* **13**, 7357. <https://doi.org/10.1038/s41467-022-34948-5> (2022).
- Shahverdi, A. *et al.* Mode selective up-conversion detection for lidar applications. *Opt. Express* **26**, 15914–15923. <https://doi.org/10.1364/OE.26.015914> (2018).
- Zhu, S. *et al.* Quantum parametric mode sorting: A case study on small angle scattering. *JOSA B* **38**, D15–D21. <https://doi.org/10.1364/JOSAB.430550> (2021).
- Lee, J. *et al.* Quantum parametric mode sorting lidar for measurement of snow properties (2022). AGU Fall Meeting.
- Timmermann, K. E. & Nowak, R. D. Multiscale modeling and estimation of Poisson processes with application to photon-limited imaging. *IEEE Trans. Inf. Theory* **45**, 846–862. <https://doi.org/10.1109/18.761328> (1999).
- Harmany, Z. T., Marcia, R. F. & Willett, R. M. This is spiral-tap: Sparse Poisson intensity reconstruction algorithms-theory and practice. *IEEE Trans. Image Process.* **21**, 1084–1096. <https://doi.org/10.1109/TIP.2011.2168410> (2012).
- Oh, A. K., Harmany, Z. T. & Willett, R. M. Logarithmic total variation regularization for cross-validation in photon-limited imaging. In *2013 IEEE International Conference on Image Processing*, 484–488 (2013).
- Heffes, H. & Lucantoni, D. M. A Markov modulated characterization of packetized voice and data traffic and related statistical multiplexer performance. *IEEE J. Select. Areas Commun.* **4**, 856–868. <https://doi.org/10.1109/JSAC.1986.1146393> (1986).

26. Snyder, D. L., Hammoud, A. M. & White, R. L. Image recovery from data acquired with a charge-coupled-device camera. *JOSA A* **10**, 1014–1023. <https://doi.org/10.1364/JOSAA.10.001014> (1993).
27. Umasuthan, M., Wallace, A., Massa, J., Buller, G. & Walker, A. Processing time-correlated single photon counting data to acquire range images. *IEEE Proc. Vis. Image Signal Process* **145**, 237–243. <https://doi.org/10.1049/ip-vis:19982152> (1998).
28. Altmann, Y., Ren, X., McCarthy, A., Buller, G. S. & McLaughlin, S. Robust bayesian target detection algorithm for depth imaging from sparse single-photon data. *IEEE Trans. Comput. Imaging* **2**, 456–467. <https://doi.org/10.1109/TCL.2016.2618323> (2016).
29. Halimi, A., Maccarone, A., McCarthy, A., McLaughlin, S. & Buller, G. S. Object depth profile and reflectivity restoration from sparse single-photon data acquired in underwater environments. *IEEE Trans. Comput. Imaging* **3**, 472–484. <https://doi.org/10.1109/TCL.2017.2669867> (2017).
30. Marais, W. J. *et al.* Approach to simultaneously denoise and invert backscatter and extinction from photon-limited atmospheric lidar observations. *Appl. Opt.* **55**, 8316–8334. <https://doi.org/10.1364/AO.55.008316> (2016).
31. Snyder, D. L. & Miller, M. I. *Random point processes in time and space* 2nd edn. (Springer-Verlag, 1991).
32. Beck, A. & Teboulle, M. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *IEEE Trans. Image Process.* **18**, 2419–2434. <https://doi.org/10.1109/TIP.2009.2028250> (2009).
33. Hayman, M., Stillwell, R. A. & Spuler, S. M. Optimization of linear signal processing in photon counting lidar using Poisson thinning. *Opt. Lett.* **45**, 5213–5216. <https://doi.org/10.1364/OL.396498> (2020).
34. Spuler, S. M. *et al.* Micropulse dial (mpd)—a diode-laser-based lidar architecture for quantitative atmospheric profiling. *Atmos. Meas. Tech.* **14**, 4593–4616. <https://doi.org/10.5194/amt-14-4593-2021> (2021).
35. NCAR/EOL Remote Sensing Facility. NCAR MPD data. UCAR/NCAR—Earth Observing Laboratory. <https://doi.org/10.26023/MX0D-Z722-M406>. Accessed 2022.
36. Computational and Information Systems Laboratory, CISL. Cheyenne: HPE/SGI ICE XA System (NCAR Community Computing). Tech. Rep., National Center for Atmospheric Research (2020).
37. Hayman, Matthew, Stillwell, Robert A, Carnes, Josh, & Spuler, Scott M. Data for 2D Signal Estimation for Sparse Distributed Target Photon Counting Data. UCAR/NCAR—Earth Observing Laboratory, <https://doi.org/10.5281/zenodo.8341854>. Accessed 2023.

Acknowledgements

We would like to acknowledge high-performance computing support from Cheyenne and Casper³⁶ provided by NSF-NCAR's Computational and Information Systems Laboratory (CISL), sponsored by the National Science Foundation. This material is based upon work supported by the National Center for Atmospheric Research, which is a major facility sponsored by the National Science Foundation under Cooperative Agreement No. 1852977, and NASA 21-NSTGRO22-0326. We would also like to acknowledge the use of Poisson Total Variation software developed by Willem Marais and his expert advice in developing the solutions described in this work.

Author contributions

M.H. developed, implemented and analyzed the results of the algorithms and simulations. J.C. Developed the data acquisition hardware for the demonstration. R.S. Integrated the data acquisition hardware with the lidar instrument and acquired the data for processing. G.K., S.S. and J.T. provided technical expertise and advice. M.H. wrote the main manuscript text. All Authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024